STATISTICAL TABULATION AND PRACTICE.*

By Roland P. Falkner, Ph.D., *Alexander Hamilton Institute.*

Nowadays the use of statistics is widespread and resort to tables or to graphic representation permeates modern literature on all subjects. There seems scarce a subject under heaven which someone does not present to us in tabular form or seek to make clear to us by means of charts and diagrams. What was at one time the peculiar reserve of the statistician and the economist has become the hunting ground of all the sciences, the pseudo sciences, and the advertiser. Statistics have become a common language. Into this language, whether in tables of figures or in graphic forms, everyone ventures boldly without even a Meisterschaft or a correspondence course preparation.

Most of us have blundered into it, and even those who may with best right claim to be experts, would admit that with them statistics is more of a habit than a principle. The result is an extreme diversity in habits of expression, which defeats the first purpose of statistical methods, whether in figures or diagrams, which is clearness of expression for groups of facts.

It is perhaps in the field of graphic presentation that the tyro indulges in the wildest flights of misrepresentation. The search for the striking and unusual leads to a perversion of fact. But when there is not conscious effort, inexperience often leads to confusing and misleading attempts to portray facts graphically. Earnest efforts are being made by the joint committee whose report was printed in the December issue of the Quarterly Publications to establish definite standards and, if they succeed in this laudable endeavor, much misdirected energy should be turned into profitable channels. Very little has been done in the search for standards in constructing and printing statistical tables.

The problems of tabular presentation of facts are comparatively simple and have not consciously been faced by

* Paper presented at the annual meeting of the American Statistical Association, Washington, D. C., December 29, 1915.

many statisticians. For that reason, they are also somewhat technical, but they are none the less real. The object of this paper is to present certain considerations in regard to arrangement and use of tables, with a view to raising the question whether standardization is necessary or advisable.

Problems of tabular presentation have had little consideration outside of government bureaus of statistics, each of which in the interest of uniformity in its publications, and wholly independent of the others, has developed rules governing the practice to be followed. There seems to be no reason why a like need should not be felt by the larger research foundations and by the statistical journals.

The ordinary printed table, perhaps not more than four or five columns wide and oftentimes not more than half as deep as the page on which it is printed, offers little difficulty. The eye takes in all its parts at a glance and the arrangement is not of vital importance. Difficulties begin only with larger tables. These we shall consider, and if rules for their construction were to be adopted, it would seem reasonable to apply such rules to smaller tables as well.

Without pretending to exhaust the infinite complexity of facts and their relations which may be expressed in tabular form, our attention will be directed to certain frequently recurring types which it is hoped will sufficiently emphasize what appear to be underlying principles of table construction.

The Whole and Its Parts, Stub Arrangement. One of the most frequent series of facts to be recorded is when facts are given for some total, say a geographic unit and its component parts, these being arranged in the stubs (Exhibit A).

The common arrangement is to name the parts first and place the total at the bottom (Table I). The reason given for such an arrangement is usually that in a sum in addition, the total is placed at the bottom. This would be conclusive if the addition were the main fact presented in the table, but it is obviously subsidiary. The chief interest attaches to the relations of the parts to the whole, and as there is one factor common to all these relations, the total, this is the most conspicuous fact to be exhibited.

EXHIBIT A.

POPULATION, AREA AND DENSITY, NEW ENGLAND, 1910.

Alternative Arrangement.

I.

Division and State	Population	Area in Sq. Miles	Population per Sq. Mile
Maine	742,371	29,895	24.8
New Hampshire	430,572	9,031	47.7
Vermont	355,956	9,124	39.0
Massachusetts	3,366,416	8,039	418.8
Rhode Island	542,610	1,067	508.5
Connecticut	1,114,756	4,820	231.3
New England	**6,552,681**	**61,976**	**105.7**

II.

Division and State	Population	Area in Sq. Miles	Population per Sq. Mile
New England	**6,552,681**	**61,976**	**105.7**
Maine	743,371	29,895	24.8
New Hampshire	430,572	9,031	47.7
Vermont	355,956	9,124	39.0
Massachusetts	3,366,416	8,039	418.8
Rhode Island	542,610	1,067	508.5
Connecticut	1,114,756	4,820	231.3

It is not infrequent that we desire to display the whole in relation to *some* of its component parts, *e. g.*, total wheat production in bushels and most important wheat producing states, without *all* parts. Here the analogy with a sum in addition fails entirely.

If in such a table, the center of interest is the total, in our illustration, New England, it cannot fail to be noted that this arrangement places the most conspicuous fact in the poorest position in the table as respects legibility. If a table extends over a page, this disadvantage is greatly emphasized. Many tables of the Statistical Abstract require the book to be turned up and fill two pages with the total at the bottom of the second page. Here a poor adjustment of the pages further complicates the situation. One table given here often covers 30 pages with the total on the last page. (Table 237, 1913, pp. 412 to 441. Exports of domestic merchandise.)

The census usage places the total at the top (Table II). This obviates all the disadvantages of the common arrangement. It places the most conspicuous fact in the posi-

tion where it is most easily read whether the table occupies two inches, a page, or several pages. It represents a finished product, not a process. Hence any conclusions drawn from the process of addition do not apply. The parts displayed need not under this arrangement equal the total. Parts may be omitted, or parts may be subdivided as for instance where in a list of states it may be desirable following New York State, to give the same facts for New York City.

The Whole and Its Parts, Box Arrangement. Convenience of arrangement not infrequently dictates placing the whole and its parts in the box heads rather than the stub (Exhibit B.)

EXHIBIT B.

PER CENT. OF TOTAL POPULATION IN CERTAIN AGE CLASSES, NEW ENGLAND, 1910.

Alternative Arrangements.

I.

Ages in Years	Me.	N. H.	Vt.	Mass.	R. I.	Conn.	New England
Under 5.....	9.7	9.2	9.6	9.8	10.0	10.1	**9.8**
5–14.....	17.7	17.0	18.0	17.2	17.6	17.7	**17.4**
15–24.....	17.1	17.4	16.8	18.5	19.5	18.8	**18.3**
25–44.....	28.2	29.2	28.5	32.5	32.0	31.7	**31.4**
45–64.....	18.9	19.2	18.8	16.7	16.2	16.3	**17.1**
67 and over.....	8.2	7.9	8.2	5.2	4.6	5.3	**5.9**

II.

Ages in Years	New England	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
Under 5.....	**9.8**	9.7	9.2	9.6	9.8	10.0	10.1
5–14.....	**17.4**	17.7	17.0	18.0	17.2	17.6	17.7
15–24.....	**18.3**	17.1	17.4	16.8	18.5	19.5	18.8
25–44.....	**31.4**	28.2	29.2	28.5	32.5	32.0	31.7
45–64.....	**17.1**	18.9	19.2	18.8	16.7	16.2	16.3
65 and over.....	**5.9**	8.2	7.9	8.2	5.2	4.6	5.3

Here again alternative arrangements are possible. Frequently the whole is placed last in the column (Table I) and this follows similar impulses which led to putting the total at the bottom of the stub. We read from left to right and if the total is regarded as the result of the parts the last column is the logical place for it. Cross addition is less frequently used than up and down addition, but when it is used, it generally proceeds from left to right.

But if we emphasize the fact that a statistical table is a result and not a process, all the weight of argument, whatever may be the usage, favors placing the total in the column following the stub (Table II). Here it can be most easily read, and as it is often the central fact around which the table revolves, it is entitled to the best place.

Sequences, Stub Arrangement. A frequent type of statistical table shows similar facts for a sequence of years or months, the facts being recorded in the boxes, the years or months in the stubs (Exhibit C).

EXHIBIT C.

DENSITY OF POPULATION, NEW ENGLAND, 1860–1910.

Different Arrangements.

I.

Year.	New England.	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
1860	**50.6**	21.0	36.1	34.5	153.1	163.7	95.5
1870	**56.3**	21.0	35.2	36.2	181.3	303.7	111.5
1880	**64.7**	21.7	38.4	36.4	221.8	259.2	129.2
1890	**75.8**	22.1	41.7	36.4	278.5	323.8	154.8
1900	**90.2**	23.2	45.6	34.7	349.0	461.6	188.5
1910	**105.7**	24.8	47.7	39.0	418.8	508.5	231.3

II.

Year.	New England.	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
1910	**105.7**	24.8	47.7	39.0	418.8	508.5	231.3
1900	**90.2**	23.2	45.6	37.7	349.0	461.6	188.5
1890	**75.8**	22.1	41.7	36.4	278.5	323.8	154.8
1880	**64.7**	21.7	38.4	36.4	221.8	259.2	129.2
1870	**56.3**	21.0	35.2	36.2	181.3	203.7	111.5
1860	**50.6**	21.0	36.1	34.5	153.1	163.7	95.5

III.

Year.	New England.	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
1910	**105.7**	**24.8**	**47.7**	**39.0**	**418.8**	**508.5**	**231.3**
1860	**50.6**	21.0	36.1	34.5	153.1	163.7	95.5
1870	**56.3**	21.0	35.2	36.2	181.3	203.7	111.5
1880	**64.7**	21.7	38.4	36.4	221.8	259.2	129.2
1890	**75.8**	22.1	41.7	36.4	278.5	323.8	154.8
1900	**90.2**	23.2	45.6	37.7	349.0	401.6	188.5
1910	**105.7**	24.8	47.7	39.0	418.8	508.5	231.3

The usual order places the earliest date at the top and the latest at the bottom. As we read a page from top to bottom,

we naturally expect a sequence in this order (Table I). While this arrangement corresponds with the normal order of a sequence it has the disadvantage of placing the last date in the most inconspicuous position.

A contrary arrangement familiar in census publications places the latest date at the top and the earliest date at the bottom (Table II). Perhaps all of us have seen men who were not figuratively, but literally as bald as a billard ball. It disconcerted us at first, but in time we got used to it. This table produces the same effect. Constant use may familiarize us, but we never quite forget the shock of first acquaintance. Its awkwardness remains and we constantly find ourselves reading down instead of up, and a conscious effort is necessary to get at the true meaning.

None the less, the census usage is not a monstrosity but well founded. It should be remembered that the primary interest in census figures is the picture which they give of present conditions and these are emphasized when the latest figures are given the first place.

The reason for such totally different practices lies in the distribution of emphasis. What is most important, the sequence or the present status? The first arrangement emphasizes the sequence, the second the present status. Both of these factors of importance enter into any figures of this character.

In this apparently insoluble conflict, it may be that the most satisfactory usage would be a compromise which repeated the latest year and gave it both the first and the last place (Table III).

Sequences, Box Arrangement. When the years or months appear in the boxes, a somewhat similar situation arises (Exhibit D).

If the normal sequence from left to right is followed (Table I), the last date comes in the last column in the least conspicuous part of the table where it is least easily read. In so far as the interest centres in present conditions as in census enumerations, this is a distinct disadvantage.

EXHIBIT D.

DENSITY OF POPULATION—NEW ENGLAND, 1860–1910.

Different Arrangements.

I.

Division and State.	1860.	1870.	1880.	1890.	1900.	1910.
New England	**50.6**	**56.3**	**64.7**	**75.8**	**90.2**	**105.7**
Maine	21.0	21.0	21.7	22.1	23.2	24.8
New Hampshire	36.1	35.2	38.4	41.7	45.6	47.7
Vermont	34.5	36.2	36.4	36.4	37.7	39.0
Massachsetts	153.1	181.3	221.8	278.5	349.0	418.8
Rhode Island	163.7	203.7	259.2	323.8	401.6	508.6
Connecticut	95.5	111.5	129.2	154.8	188.5	231.3

II.

Division and State.	1910.	1900.	1890.	1880.	1870.	1860.
New England	**105.7**	**90.2**	**75.8**	**64.7**	**56.3**	**50.6**
Maine	24.8	23.2	22.1	21.7	21.0	21.0
New Hampshire	47.7	45.6	41.7	38.4	35.2	36.1
Vermont	39.0	37.7	36.4	36.4	36.2	34.5
Massachusetts	418.8	349.0	278.5	221.8	181.3	153.1
Rhode Island	508.5	401.6	323.8	259.2	203.7	163.7
Connecticut	231.3	188.5	154.8	129.2	111.5	95.5

III.

Division and State.	1910.	1860.	1870.	1880.	1890.	1900.	1910.
New England	**105.7**	**50.6**	**56.3**	**64.7**	**75.8**	**90.2**	**105.7**
Maine	**24.8**	21.0	21.0	21.7	22.1	23.2	24.8
New Hampshire	**47.7**	36.1	35.2	28.4	41.7	45.6	47.7
Vermont	**39.0**	34.5	36.2	36.4	36.4	37.7	39.0
Massachusetts	**418.8**	153.1	181.3	221.8	278.5	349.0	418.8
Rhode Island	**508.5**	163.7	203.7	259.2	323.8	401.6	508.5
Connecticut	**231.3**	95.5	111.5	129.2	154.8	188.5	231.3

On the other hand, the arrangement of dates so that the latest immediately follows the stub and the earliest comes in the last column (Table II) as in the census publications, offends our natural instinct to read from left to right and interferes with the ease with which the sequence is comprehended.

Here again a compromise between conflicting interests may be suggested by giving first the latest date, giving earlier ones in the natural order, and repeating for the sake of the sequence the last date in the last column (Table III).

Mechanical Aids to Clearness. The object of the statistical table is to compress into a small space a body of information the narration of which in the text would be cumbersome and exhausting to the reader. It is in short a method of condensation, and it is of the utmost importance that, as it tells so much in so small a compass, it tell it as clearly as practicable. Statisticians cannot and others should not, therefore, neglect the various mechanical aids to clearness which are available. These are chiefly differences in type and rules. The total should always be distinguished from the parts either by the use of heavier type, or by a rule or space separating it from the parts, or by both methods. When the total is in the box head, a double rule or heavy rule is advisable.

In a chronological series emphasis can be placed upon the latest date by similar methods and this is often times desirable.

Numbers and Titles for Tables. The usual practice is to give each table a number and no other method affords such a simple plan for identification in the text comment. It has sometimes been argued that numbering tables makes them a thing apart from the text instead of part and parcel of the treatment of subject. Without contesting the plausibility of this argument, it must be admitted that a table of figures is just as much and no more a part of the text than is a map in any geographical publication. It is somewhat less distinct perhaps than an illustration, but a certain separateness cannot be denied and cannot be avoided by ignoring it. The omission of table numbers leads to difficulties in practice which it is the first duty of the statistician to avoid.

Whether tables should have titles as well as numbers is largely a question of taste. Larger tables commonly have such titles, but they are often omitted in text tables. If the title is helpful in text references, it should not be omitted. My personal preference is for the use of titles in every instance no matter how large or how small the table. In any case, long titles are to be avoided. The practice of some government bureaus of giving in the title all the elements which enter into the box heads, involves useless repetition. When, for any reason, long descriptive titles are felt to be necessary, a compromise may be effected by the use of a short catch title, followed by the descriptive title.

Contents of Tables, Abbreviated Statements. The contents of statistical tables vary so much according to the nature of the problems involved, that it might seem that we should look in vain for any general rules. However, one type of problem constantly recurs wherever round numbers or percentages are used. Statistical accuracy is often served by giving such larger items as wheat production, value of farms, manufactured products in terms of millions of dollars. Yet our practice in such cases is to grasp after even greater exactness. We may give one decimal after the millions as notice perhaps that the numbers are not round after all. Then we add up the abbreviated figures and carefully adjust the decimals to see that they add up exactly to the sum given in the detailed figures.

Again we frequently have a division of some total into percentages. These again are often given with one or two decimals attached. Again we are obsessed with the idea that the sum of such percentages should equal 100 and we painfully "adjust" them to bring about such a result.

In any proceeding of this kind, we are obliged in some measure to twist facts to make them suit the table. If we were to really grasp the idea that in presenting a statistical table, we are giving not a sum in addition but a finished result, we should not expect our tables to prove. Whether the sum of our percentages equalled 99.7 or 100.5, each item would be correctly stated and we should have done no violation to our conscience in doctoring the figures to get the desired result.

Conclusion. A few of the questions which arise in the practice of statistics have been discussed in this paper. To those who are not at least in some degree versed in the subject, they might appear trivial and inconsequential. But to those who are interested in perfecting the precision of their instruments of thought and expression, they have their significance. If the aim of all statistical statements is clearness and sanity we cannot neglect anything which contributes to this end.